

Contents

List of Figures	iii
List of Tables	iv
1 Introduction	1
1.1 The DANCe project	1
2 The Pleiades as a benchmark	3
2.1 Generalities	3
2.2 Spatial Distribution	3
2.3 Velocity Distribution	3
2.4 Luminosity Distribution	3
2.5 Mass Distribution	3
2.6 The current dynamical scenario	3
2.7 The Pleiades DANCe DR2.	3
3 Bayesian formalism	4
3.1 Introduction to probability theory.	4
3.2 Bayesian Hierarchical Models	9
3.3 Modelling the data	9
3.4 Priors	9
3.5 Sampling the posterior distribution	9
3.6 Codes	10
4 Results	11
4.1 Performance of the classifier	11
4.2 Velocity distribution	11

	ii
4.3 Spatial distribution	11
4.4 Luminosity distribution	11
4.5 Mass distribution	11
4.6 The mass distribution on time	11
4.7 The phase space	11
4.8 Updating the previous knowledge	11
5 Conclusions and Future Work	12
A Appendix Chapter Title	13

List of Figures

List of Tables

Chapter 1

Introduction

Description of:

- Motivation to investigate nearby young open clusters.
- The importance of the IMF
- Current state of knowledge in dynamical simulations.
- The problematic of constraints in dynamical theories.

It must be clear which is the objective

Description of the Nearby open clusters and their properties.

1.1 The DANCe project

- List of open clusters in the DANCe project.
- the importance of the pleiades, why we restrict to it.

It must be clear what are the limitations, the boundaries in which the objective will be searched

Description of the current methodologies used to address the question mentioned previously.

- The works of Sarro, Krone-Martins, Malo, Gagne etc.
- The advantages and caveats of the previous methodologies.
- It must be clear the necessity of a new perspective

The proposal we made. The use of Bayesian Hierarchical Models. Benefits and issues of BHM.

Description of the advantages of BHM.

-¿ It must be clear that BHM are the best choice.

Description of the practical issues needed to be solved in order to use BHM.

MCMC techniques and PSO.

-¿ It must be clear that MCMC methods are the best option.

Brief descriptions of our results and how they impact our current knowledge.

-¿ It must be clear that we attained the objective: The pleiades velocity, spatial and mass distributions.

Chapter 2

The Pleiades as a benchmark

2.1 Generalities

Description of the current knowledge on the pleiades cluster Age, distance, metallicity reddening, number of members,

2.1.1 The distance controversy

2.2 Spatial Distribution

2.3 Velocity Distribution

2.3.1 Radial Velocity

2.4 Luminosity Distribution

2.5 Mass Distribution

2.6 The current dynamical scenario

2.7 The Pleiades DANCe DR2.

Chapter 3

Bayesian formalism

This chapter provides a general introduction to probability theory and its application to parametric inference. Since the objective of this work is to infer the probability distributions of the cluster properties (e.g. luminosity and velocity), I give reason to prove that the Bayes' theorem provides the proper framework for the inference of the parameters governing these distributions. Later in this chapter, I describe the reason for which the Bayesian Hierarchical Models are the best option to parametric inference under the Bayesian framework.

In the following Sections I will describe in detail the assumptions I made to model the data and to select the prior distributions. The two final Sections of this Chapter focus on: the practical issues related to the sampling of the cluster distributions, and the description of details and assumptions embed in the codes I developed.

3.1 Introduction to probability theory.

Uncertainty and probability are closely entangled. Anything we measure has an associated uncertainty, otherwise is not a complete measurement¹. The term uncertainty must not be confused with the term error, which refers to the difference between the measured value of the quantity and the *true* value² of it [?]. It is commonly agreed that uncertainty of a measurement can be expressed in a

¹Upper and lower limits are examples of incomplete measurements.

²The true value is that which ideally results when the uncertainty tends to zero.

probabilistic basis [?]. This means that whenever we measure a quantity, lets say a , the distribution of the repeated measurements of a , is a probability distribution function, $p(a)$. As any other probability distribution, $p(a)$ satisfies the following properties:

Property 1 It has units, those of the inverse of a .

Property 2 $p(a) \geq 0$.

Property 3 $1 = \int_a p(a) da$.

These properties hold regardless of the dimension of a , it means that the joint uncertainty of all measured quantities of an object is also a probability distribution. Furthermore, they also hold if the probability distribution is conditioned in any other quantity. Lets imagine that we measure the positions, projected in the plane of sky (the plane perpendicular to the line of sight), of one star, these measurements are conditioned in the magnitude (brightness) of the object we measure. If the object is too bright, like the sun, it will saturate the detector and it will render the measurement useless. On the other hand, if the object is too faint we simple will not have enough photons to measure it. So, the stellar positions in the sky, which we can call a and b because they are two dimensions, are conditioned on the magnitude, c , of the object. Therefore, $p(a, b|c)$ must also satisfy:

- It has units of $a^{-1}b^{-1}$.
- $p(a, b|c) \geq 0$.
- $1 = \int_a \int_b p(a, b|c) da \cdot db$.

The link between joint and conditioned probabilities is given by the following symmetric definition:

$$\begin{aligned} p(a, b) &= p(a|b) \cdot p(b). \\ p(a, b) &= p(b|a) \cdot p(a). \end{aligned} \tag{3.1}$$

This can be further conditioned on c to obtain:

$$\begin{aligned} p(a, b|c) &= p(a|b, c) \cdot p(b|c), \\ p(a, b|c) &= p(b|a, c) \cdot p(a|c), \end{aligned} \tag{3.2}$$

If the joint probability of a and b can be factorised, this is

$$\begin{aligned} p(a, b) &= p(a) \cdot p(b), \\ p(a, b) &= p(b) \cdot p(a), \end{aligned} \tag{3.3}$$

then a and b are say to be *independent*. An alternative option is to say that a and b are *independent*, if the conditional probability of a on b is $p(a|b) = p(a)$.

The most important thing we can do with probability distributions is to integrate them. **Property 3** establish that the amount³ of probability $p(a)$ spread over the volume of the support of a adds to one. This Property allows us to *marginalise* any non-desired variable. Lets imagine again that a and b are the measured positions of some star and we have several measurements of these positions. Then we will have the joint probability distribution of them, $p(a, b)$ (must likely it will be a bivariate gaussian but that does not matter now). If we are interested lets say in the mean value of a , we first must get rid of b . For it, we *marginalise* out b in the following way,

$$p(a) = \int_b p(a, b) \cdot db. \tag{3.4}$$

Then we compute the *expected value* of a , $E(a)$, which is identified with the mean of a once we have drawn many realisations from its probability distribution. To compute it, we add all the possible values of a weighted by their probability. This is,

$$E(a) = \int_a a \cdot p(a) \cdot da. \tag{3.5}$$

³Which could be infinite, like in Dirac's delta.

Once again, these last two equations (3.4 and 3.5) hold in case they are conditioned in any other measurement. For example, the magnitude of the object, as in our previous analogy. Notice however that, once the brightness of the object lay within in the dynamic range of the detector, a and b became *independent* of the magnitude.

It is important to recall that the term measurement, and its unavoidable uncertainty, refer not just to directly measured quantities, like the photons (counts) and pixels in a CCD, but also to indirect measurements. Stellar magnitudes and positions in the sky, for example, are indirect measurements derived from the direct measurement of photons, pixels and telescope arrangement. This generalisation also applies to the measurement of parameters in any physical or statistical model, like the one I will describe in Section ??.

3.1.1 Bayes theorem

The definition of conditioned probability (Eq. 3.2) leads to the Bayes' theorem:

$$p(a|b, c) = \frac{p(b|a, c) \cdot p(a|c)}{p(b|c)}. \quad (3.6)$$

Integrating on a we find that,

$$\begin{aligned} p(b|c) \cdot \int_a p(a|b, c) \cdot da &= \int_a p(b|a, c) \cdot p(a|c) \cdot da \\ Z \equiv p(b|c) &= \int_a p(b|a, c) \cdot p(a|c) \cdot da. \end{aligned} \quad (3.7)$$

In this last equation Z refers to what is known as the *evidence*. I will come back to this *evidence* in a few paragraphs. This last Eq. also illustrates that $p(b|c)$ is a normalisation constant which can be evaluated once $p(b|a, c)$ and $p(a|c)$ are known. These two terms are commonly referred as the *likelihood* of the data ($p(b|a, c)$), and the *prior* ($p(a|c)$). These names arise in the context of parametric inference, where a, b and c are identified, respectively, with the parameters in a model, the data which the model tries to describe, and the information used in the construction of the model. Thus, the term $p(a|b, c)$ is called the *posterior* distribution of the parameters given the data and of course it is also conditioned in the model and the

information used to construct it. Notice that, although formally the likelihood and the prior are probability distributions in b and a , respectively, for $p(a|b, c)$ to be a probability distribution on a , it only suffices that the product of the likelihood times the prior does not vanish everywhere or be negative anywhere⁴. In this case, they are called *improper* priors or likelihoods. If their product vanishes everywhere, which may be the case if the prior is terribly specified or if the likelihood does not take proper account of extreme data, then the posterior is not a probability distribution due to a division by zero. In any case, it makes no sense try to estimate the parameters of a model with zero evidence.

The Bayes' theorem can be interpreted as the probabilistic way to update knowledge. In other words, the way to update knowledge when we recognise the uncertainty associated to it. When I say knowledge I refer to abstractions or models that we made about nature. In my perspective, they are always uncertain, even if its uncertainty is negligible given the current evidence that support them.

Whenever we have a model, we have prior knowledge over it. Actually, this can be classified in to kinds of priors. One refers to the prior information conveyed in the model, which I call M . It is the information that the creator of the model use to establish the relations among the elements of the model: data and parameters. The second kind of prior, which coincides with the previous use of the term, refers to the statement the user of the model made of his/her believes about the distribution of the parameter values. This is indeed subjective. However, it is, in my opinion, less subjective than the former, M , prior information. At least in this last kind, the subjectivity is expressed objectively in a probabilistic, and therefore measurable way.

That said, the Bayes' theorem helps us to update our prior believes by means of the data, once we multiply them by the likelihood of it. Then, the posterior probabilities of the parameters given the current data, became the new prior believes once more data is available. Furthermore, the Bayes' theorem also provides the objective way to compare two models and update the prior information, M , used to construct them. This is called model selection.

⁴Although negative probabilities may have sense in quantum mechanics. See for example [?]]

3.1.2 Model Selection

Whenever we have a collection of models, even two of them, and a data set, the most frequent thing we do is try to compare them with each other. Most of the time, if not always, when we fit a collection of models is to find the one that best describes the data.

3.1.3 Membership probability

3.2 Bayesian Hierarchical Models

3.2.1 Generalities

Comparisons with other techniques in IA. It must be clear the BHM are the only option.

3.2.2 Examples

Its applications in IA. Its applications in astrophysics.

3.2.3 Graphical representation: Probabilistic Graphical Models

3.3 Modelling the data

Positions, proper motions, photometry

3.3.1 Missing values

3.3.1.1 Missing value pattern

3.3.2 The field population

3.3.3 The cluster population

3.4 Priors

3.5 Sampling the posterior distribution

Description of the techniques used to obtain samples from the posterior distribution.

History and used versions.

3.5.1 PSO

3.5.1.1 The charged PSO

3.5.2 MCMC

3.5.2.1 Generalities

3.5.2.2 Flavours

HMC,NUTS,Gibbs, Metropolis-Hasting, Affine invariant, stretch-move, Multi-Nest. It must be clear why we choose emcee and multinest

3.5.2.3 Convergence

3.5.2.4 The evidence

3.6 Codes

3.6.0.5 The modified charged PSO

3.6.0.6 Improvements of emcee

3.6.0.7 The GMM with missing values

3.6.1 Parallel implementations

Description of the implementation. MPI, python stan, etc. explain in detail the difficulties faced at implementing the different codes in the different servers.

Chapter 4

Results

4.1 Performance of the classifier

Compare and explain the differences with Stauffer members.

4.2 Velocity distribution

4.3 Spatial distribution

4.4 Luminosity distribution

4.5 Mass distribution

4.5.1 The mass-luminosity relation

4.6 The mass distribution on time

4.7 The phase space

4.8 Updating the previous knowledge

Chapter 5

Conclusions and Future Work

Appendix A

Appendix Chapter Title

In the following we include